

**MEPS HC-122:
Panel 12 Longitudinal Data File
December 2010**

**Agency for Healthcare Research and Quality
Center for Financing, Access, and Cost Trends
540 Gaither Road
Rockville, MD 20850
(301) 427-1406**

Table of Contents

A.	Data Use Agreement	A-1
B.	Background	B-1
	1.0 Household Component.....	B-1
	2.0 Medical Provider Component	B-1
	3.0 Survey Management and Data Collection	B-2
C.	Technical and Programming Information	C-1
	1.0 General Information.....	C-1
	2.0 Data File Information.....	C-1
	2.1 Variables	C-2
	2.1.1 Variables from Annual Full-Year Consolidated Files.....	C-2
	2.1.2 Constructed Variables for Selection of Analytic Group Files.....	C-6
	2.1.3 Estimation Variables	C-7
	2.1.4 Prescription Medications Use and Expenditures and Total Expenditures	C-8

A. Data Use Agreement

Individual identifiers have been removed from the micro-data contained in these files. Nevertheless, under sections 308 (d) and 903 (c) of the Public Health Service Act (42 U.S.C. 242m and 42 U.S.C. 299 a-1), data collected by the Agency for Healthcare Research and Quality (AHRQ) and/or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which they were supplied; any effort to determine the identity of any reported cases is prohibited by law.

Therefore in accordance with the above referenced Federal Statute, it is understood that:

1. No one is to use the data in this data set in any way except for statistical reporting and analysis; and
2. If the identity of any person or establishment should be discovered inadvertently, then (a) no use will be made of this knowledge, (b) the Director Office of Management AHRQ will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHRQ, and (d) no one else will be informed of the discovered identity; and
3. No one will attempt to link this data set with individually identifiable records from any data sets other than the Medical Expenditure Panel Survey or the National Health Interview Survey.

By using these data you signify your agreement to comply with the above stated statutorily based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates Title 18 part 1 Chapter 47 Section 1001 and is punishable by a fine of up to \$10,000 or up to 5 years in prison.

The Agency for Healthcare Research and Quality requests that users cite AHRQ and the Medical Expenditure Panel Survey as the data source in any publications or research based upon these data.

B. Background

1.0 Household Component

The Medical Expenditure Panel Survey (MEPS) provides nationally representative estimates of health care use, expenditures, sources of payment, and health insurance coverage for the U.S. civilian non-institutionalized population. The MEPS Household Component (HC) also provides estimates of respondents' health status, demographic and socio-economic characteristics, employment, access to care, and satisfaction with health care. Estimates can be produced for individuals, families, and selected population subgroups. The panel design of the survey, which includes 5 Rounds of interviews covering 2 full calendar years, provides data for examining person level changes in selected variables such as expenditures, health insurance coverage, and health status. Using computer assisted personal interviewing (CAPI) technology, information about each household member is collected, and the survey builds on this information from interview to interview. All data for a sampled household are reported by a single household respondent.

The MEPS-HC was initiated in 1996. Each year a new panel of sample households is selected. Because the data collected are comparable to those from earlier medical expenditure surveys conducted in 1977 and 1987, it is possible to analyze long-term trends. Each annual MEPS-HC sample size is about 15,000 households. Data can be analyzed at either the person or event level. Data must be weighted to produce national estimates.

The set of households selected for each panel of the MEPS HC is a subsample of households participating in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics. The NHIS sampling frame provides a nationally representative sample of the U.S. civilian noninstitutionalized population and reflects an oversample of blacks and Hispanics. In 2006, the NHIS implemented a new sample design, which included Asian persons in addition to households with black and Hispanic persons in the oversampling of minority populations. MEPS further oversamples additional policy relevant sub-groups such as low income households. The linkage of the MEPS to the previous year's NHIS provides additional data for longitudinal analytic purposes.

2.0 Medical Provider Component

Upon completion of the household CAPI interview and obtaining permission from the household survey respondents, a sample of medical providers are contacted by telephone to obtain information that household respondents can not accurately provide. This part of the MEPS is called the Medical Provider Component (MPC) and information is collected on dates of visit, diagnosis and procedure codes, charges and payments. The Pharmacy Component (PC), a subcomponent of the MPC, does not collect charges or diagnosis and procedure codes but does collect drug detail information, including National Drug Code

(NDC) and medicine name, as well as date filled and sources and amounts of payment. The MPC is not designed to yield national estimates. It is primarily used as an imputation source to supplement/replace household reported expenditure information.

3.0 Survey Management and Data Collection

MEPS HC and MPC data are collected under the authority of the Public Health Service Act. Data are collected under contract with Westat, Inc. Data sets and summary statistics are edited and published in accordance with the confidentiality provisions of the Public Health Service Act and the Privacy Act. The National Center for Health statistics (NCHS) provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports, micro data files, and tables via the MEPS web site: www.meps.ahrq.gov. Selected data can be analyzed through MEPSnet, an on-line interactive tool designed to give data users the capability to statistically analyze MEPS data in a menu-driven environment.

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Financing Access and Cost Trends, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850 (301-427-1406).

C. Technical and Programming Information

1.0 General Information

For MEPS Panels 1-8, longitudinal weight files that were released contained a limited number of variables that could be merged with data from two consecutive full-year consolidated files to create a longitudinal file for analysis. Beginning with Panel 9, AHRQ has replaced the longitudinal weight files with more complete and analytically useful panel-specific files that contain the variables from the consolidated full-year files.

This documentation describes the Panel 12 longitudinal data file from the Medical Expenditure Panel Survey Household Component (MEPS-HC). Released as an ASCII file (with related SAS and SPSS programming statements and data use information) and a SAS transport dataset, this public use file provides information collected on a nationally representative sample of the civilian noninstitutionalized population of the United States for the two-year period 2007-08. The file contains 3,438 variables and has a logical record length of 9,753 with an additional 2-byte carriage return/line feed at the end of each record.

This file consists of MEPS survey data obtained in Rounds 1-5 of MEPS Panel 12 and can be used to analyze changes over a two-year period. Variables in the file pertaining to survey administration, demographics, employment, health status, disability days, quality of care, patient satisfaction, health insurance and medical care use and expenditures were obtained from the MEPS 2007 and 2008 Full-Year Consolidated Files (HC-113 and HC-121, respectively).

The following documentation offers a brief overview of the contents and structure of the files and programming information. A codebook of all the variables included in the Panel 12 data file is provided in a separate file (H122CB.PDF). A database of all MEPS products released to date and a variable locator indicating the major MEPS data items on public use files that have been released to date can be found on the MEPS Web site: www.meps.ahrq.gov.

2.0 Data File Information

This public use file contains records for 12,440 persons in Panel 12 who were respondents for the period they were in-scope for the survey (i.e., a member of the civilian non-institutionalized population) during the two-year period. Data are available for all five rounds for 91% of the cases (11,348). The remaining 9% (1,092 persons) do not have data for one or more rounds but were in-scope for all rounds they participated in the survey. These persons are those who were born, died, were in the military or an institution, or left the country during the two-year period. In contrast, persons in the panel who participated

in the survey for only part of the period they were in-scope are not included in this file. To compensate for this attrition, adjustments were made in the construction of the panel weight variable included in this file (LONGWT). The codebook provides both weighted and unweighted frequencies for each variable on the data file. The LONGWT variable should be used to produce national estimates for the two-year period.

Each MEPS panel can be linked back to the previous years National Health Interview Survey public use data files. For information on obtaining MEPS/NHIS link files please see http://www.meps.ahrq.gov/mepsweb/data_stats/more_info_download_data_files.jsp.

2.1 Variables

2.1.1 Variables from Annual Full-year Consolidated Files

Most variables on this file were obtained from the MEPS 2007 and 2008 Full-Year Consolidated Files (HC-113 and HC-121, respectively). However, names for time dependent variables from these files were modified in order to: 1) eliminate duplicate variable names for data reflecting different time periods during the panel, and 2) standardize variable names to facilitate pooling of multiple MEPS panels for analysis.¹ Generally, annual variables with a suffix of “07” and “08” are renamed with a suffix of “Y1” and “Y2”, respectively. Variables with a suffix of “31”, “42”, and “53” are renamed with a suffix denoting the round the data was collected (i.e., “1”, “2” or “3” for variables originating from Rounds 1-3 on the 2007 full-year file and “3”, “4”, or “5” for variables originating from Rounds 3-5 on the 2008 full-year file).² It is necessary to use this crosswalk in conjunction with documentation for the 2007 and 2008 full-year consolidated files to obtain a full description of variables on this file. Table 1 below provides the crosswalk summarizing the scheme used for renaming variables from the annual files.

Table 1: Crosswalk of Variable Names between the Full-Year Consolidated files and the Longitudinal File

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Constant	No suffixes	No suffixes	<i>All variables:</i> DOBMM=DOBMM

¹ A variable named PANEL is also included to facilitate pooling across panels. This variable is simply the panel number and is therefore constant across all records within a longitudinal file.

² While round 3 values were obtained for most observations from the 2008 Full Year Consolidated File, they were obtained from the 2007 Full Year Consolidated File for sample persons where YEARIND=2 (i.e., in 2007 only).

(i.e., not round or year specific)			DOBYY=DOBYY DUID=DUID PID=PID DUPERSID=DUPERSID EDUCYR=EDUCYR HIDEG=HIDEG HISPANX=HISPANX HISPCAT=HISPCAT INTVLANG=INTVLANG RACEAX=RACEAX RACEBX=RACEBX RACEWX=RACEWX RACEX=RACEX RACETHNX=RACETHNX SEX=SEX VARPSU=VARPSU VARSTR=VARSTR
Annual, family related variables	YR	Y1 or YR1 Y2 or YR2	<i>All variables:</i> FAMIDYR=FAMIDYR1 (2007 file) FAMRFPYR=FAMRFPY1 (2007 file) FAMSZEYR=FAMSZEY1 (2007 file) FAMIDYR=FAMIDYR2 (2008 file) FAMRFPYR=FAMRFPY2 (2008 file) FAMSZEYR=FAMSZEY2 (2008 file)
Annual, CPS family identifiers	No suffix	Y1 Y2	<i>All variables:</i> CPSFAMID= CPSFAMY1 (2007) CPSFAMID= CPSFAMY2 (2008)
Annual, health insurance eligibility units	No suffix	Y1 Y2	<i>All variables:</i> HIEUIDX=HIEUIDY1 (2007) HIEUIDX=HIEUIDY2 (2008)
Annual, inscope variables	No suffixes	YR1 YR2	<i>All variables:</i> INSCOPE=INSCPYPY1 (2007 file) INSCOPE=INSCPYPY2 (2008 file)
12/31 status variables	1231 in 2007 file	Y1	<i>All variables:</i> FAMS1231=FAMSY1 (2007 file) FCRP1231=FCRPY1 (2007 file) FCSZ1231=FCSZY1 (2007 file)

	1231 in 2008 file	Y2	FMRS1231= FMRSY1 (2007 file) INSC1231=INSCY1 (2007 file) FAMS1231=FAMSY2 (2008 file) FCRP1231=FCRPY2 (2008 file) FCSZ1231= FCSZY2 (2008 file) FMRS1231= FMRSY2 (2008 file) INSC1231=INSCY2 (2008 file)
Annual	07, 07X, 07F, or 07C 08, 08X, 08F, or 08C	Y1, Y1X, Y1F, or Y1C Y2, Y2X, Y2F, or Y2C	<i>Examples:</i> TOTEXP07=TOTEXPY1 (2007 file) AGE07X=AGEY1X TOTEXP08=TOTEXPY2 (2008 file) AGE08X=AGEY2X
Variables for health insurance prior to January 1, 2007 (data collected in round 1 only)	No suffixes	No suffixes	<i>All variables:</i> PREVCOVR=PREVCOVR COVRMM=COVRMM COVRYX=COVRYX WASESTB=WASESTB WASMCARE=WASMCARE WASMCAID=WASMCAID WASCHAMP=WASCHAMP WASVA=WASVA WASPRIV=WASPRIV WASOTGOV=WASOTGOV WASAFDC=WASAFDC WASSSI=WASSSI WASSTAT1=WASSTAT1 WASSTAT2=WASSTAT2 WASSTAT3=WASSTAT3 WASSTAT4=WASSTAT4 WASOTHER=WASOTHER NOINSBEF=NOINSBEF NOINSTM=NOINSTM NOINUNIT=NOINUNIT MORECOVR=MORECOVR INSENDMM=INSENDMM INSENDYY=INSENDYY
Annual	No suffixes ³	Y1 Y2	<i>All variables:</i> KEYNESS=KEYNESY1 (2007 file) SAQELIG=SAQELIY1 (2007 file) EVRWRK=EVRWRKY1 (2008 file) EVRETIRE=EVRETIY1 (2007 file)

³To maintain the 8-character naming convention, some variable names had the last character or two dropped in the renaming process.

			EVRUNAT=EVRUNAY1 (2007 file) EVRUNINS=EVRUINY1 (2007 file) KEYNESS=KEYNESY2 (2008 file) SAQELIG=SAQELIY2 (2008 file) EVRWRK=EVRWRKY2 (2008 file) EVRETIRE=EVRETIY2 (2008 file) EVRUNAT=EVRUNAY2 (2008 file) EVRUNINS=EVRUINY2 (2008file)
Monthly	2-character month + 07 2-character month + 08	2-character month + Y1 2-character month + Y2	<i>Example:</i> PRIJA07=PRIJAY1 (2007 file) PRIJA08=PRIJAY2 (2008 file)
Round Specific	31 or 31X in 2007 file 42 or 42X in 2007 file 53 or 53X in 2007 file 31 or 31X in 2008 file 42 or 42X in 2008 file 53 or 53X in 2008 file	1 or 1X for 2007 2 or 2X for 2007 3 or 3X for 2007 3 or 3X for 2008 4 or 4X for 2008 5 or 5X for 2008	<i>Example:</i> RTHLTH31 = RTHLTH1 (2007 file) RTHLTH42 =RTHLTH2 (2007 file) RTHLTH53 =RTHLTH3 (2007 file if YRIND=2) RTHLTH31 = RTHLTH3 (2008 file if YEARIND=1 or 3) RTHLTH42 =RTHLTH4 (2008 file) RTHLTH53 =RTHLTH5 (2007 file)

Diabetes preventive care ⁴	0653, 0753, and 0853 in 2007 file	Y0R3 for 2006 data Y1R3 for 2007 data Y2R3 for 2008 data	<i>Example:</i> DSEB0653=DSEBY0R3 (2007 file) DSEY0653=DSEYY0R3 (2007 file) DSEY0753=DSEYY1R3 (2007 file) DSEY0853=DSEYY2R3 (2007 file)
	0753, 0853, and 0953 in 2008 file	Y1R5 for 2007 data Y2R5 for 2008 data Y3R5 for 2009 data	DSEB0753=DSEBY1R5 (2008 file) DSEY0753=DSEYY1R5 (2008 file) DSEY0853=DSEYY2R5 (2008 file) DSEY0953=DSEYY3R5 (2008 file)
Job Change	3142 4253	12 for 2007 23 for 2007 34 for 2008 45 for 2008	<i>All cases:</i> CHJ3142=CHJ12(2007 file) CHJ4253=CHGJ23(2007 file) YCHJ3142=YCHJ12(2007 file) YCHJ4253=YCHGJ23(2007 file) CHJ3142=CHGJ34 (2008 file) CHJ4253=CHGJ45 (2008 file) YCHJ3142=YCHGJ34 (2008 file) YCHJ4253=YCHGJ45 (2008 file)

2.1.2. Constructed Variables for Selection of Analytic Group

The following eight variables were constructed and included on the file to facilitate the selection of appropriate cases for various analyses. Table 2 below contains descriptive statistics for these variables.

YEARIND	1=both years, 2=in 2007 only, and 3=in 2008 only
ALL5RDS	Inscope and data collected in all 5 rounds (0=no, 1=yes)
DIED	Died during the two-year survey period (0=no, 1=yes)
INST	Institutionalized for some time during the two-year survey period (0=no, 1=yes)
MILITARY	Active duty military for some time during the two-year survey period (0=no, 1=yes)
ENTRSRVY	Entered survey after beginning of panel (mainly births; also includes persons who had no initial chance of selection who moved into a MEPS sample household) (0=no, 1=yes)
LEFTUS	Moved out of the country after beginning of panel (0=no, 1=yes)
OTHER	Not identified in any of the above analytic groups (0=no, 1=yes)

⁴ Diabetic foot exams, lipid profiles, and flu shots starting in 2008.

Table 2: Frequencies and Percentage for Constructed Variables

Variable	Number of Records	Percentage of Records (N=12,440)
YEARIND=1 (i.e., person in both years)	12,111	97.4
ALL5RDS=1 (yes)	11,348	91.2
DIED=1 (yes)	139	1.1
INST=1 (yes)	43	0.4
MILITARY=1 (yes)	28	0.2
ENTRSRVY=1 (yes)	780	6.3
LEFTUS=1 (yes)	53	0.4
OTHER=1 (yes)	64	0.5

Following are examples of situations where these variables would be useful in selecting records for analysis:

- Analysts interested in working only with persons who were in-scope and had data for all five rounds of the panel should subset to cases where ALL5RDS=1.
- If a researcher wanted to include persons who were in-scope and had data for all five rounds of the panel as well as those in the survey at the beginning of the panel who subsequently died, then they would include cases where ALL5RDS=1 or (ENTRSRVY=0 and DIED=1).
- If a researcher wanted to include persons who were in-scope and had data for all five rounds of the panel as well as those who died in the second year of the panel, then they would include cases where ALL5RDS=1 or (DIED=1 and YEARIND=1).

2.1.3 Estimation Variables

Longitudinal Estimations for Panel 12

The file contains a weight variable (LONGWT) and variance estimation variables (VARSTR, VARPSU) that should be applied when producing national estimates for longitudinal analyses. For example, LONGWT applied to the 11,348 cases where ALL5RDS=1 produces a weighted population estimate of 279.6 million. This represents an estimate of the number of persons in the civilian noninstitutionalized population for the entire two-year period from 2007-08. To obtain estimates of variability (such as the standard error of sample estimates or corresponding confidence intervals) for estimates based on MEPS survey data, one needs to take into account the complex sample design of MEPS by specifying the estimation variables including stratum of sample selection (VARSTR), primary sampling unit (VARPSU) and longitudinal weight (LONGWT).

Pooled Estimations

When analyzing subpopulations and/or low prevalence events, it may be desirable to pool together more than one panel of MEPS-HC data to yield sample sizes large enough to generate reliable estimates. If only data from Panels 7 and beyond are being pooled, then simply use the strata and psu variables provided on the longitudinal files for pooled estimation⁵. However, because Panels 1-6 MEPS longitudinal weight files were released with panel-specific variance structures, it is necessary to obtain the set of appropriate variance estimation variables from the HC-036 Pooled Estimation File when pooling involves these panels. This Panel 12 file also includes the set of variance estimation variables (STRA9608, PSU9608) that should be applied when producing estimates using any of the first six MEPS panels. STRA9608 and PSU9608 reconcile the differences in the variance units between the units on the released annual MEPS public use files (see HC-036 file documentation for more information).

2.1.4 Prescription Medication Use and Expenditures and Total Expenditures

This section describes an editing change in the prescription drug variables and differences in variable values between the 2007 full year consolidated file and this longitudinal file for some persons. Nearly all users of this file can simply use the summary use and expenditure variables on this file. For users who augment the longitudinal file with information from the prescription medicine files, this section provides additional information about how to account for some fills that are in both the 2007 and 2008 Prescription Medicine files.

In the third interview, where the reference period typically spans the later part of one year and the early part of the next, for each drug, the household respondent is asked to report both the number of times the drug was obtained since the last interview and the number of times the drug was obtained in the current year. When this information is missing, the total number of fills for the drug is allocated to each year (2007 or 2008). Starting with the 2008 prescription medicine data, improvements were made in the allocation method. The new method tends to allocate more of the total round 3 fills to the second year and fewer to the first year. Implementing this reallocation of fills resulted in a one-time problem for Panel 12: some fills are represented in both the 2007 and 2008 the prescription drug files. This duplication was removed from the values of the prescription drug use and expenditures for year 1 (RXTOTY1, RXEXPY1, RXSLFY1, RXMCRY1, RXMCDY1, RXPRVY1, RXVAY1, RXTRİY1, RXOFDY1, RXSTLY1, RXWCPY1, RXOSRY1, RXOPRY1, RXOPUY1, RXPTRY1, RXOTHY1) and total expenditures for year 1 (TOTEXPY1, TOTSLFY1, TOTMCRY1, TOTMCDY1, TOTPRVY1, TOTVAY1, TOTTRIY1, TOTOFDY1, TOTSTLY1, TOTWCPY1, TOTOSRY1, TOTOPRY1,

⁵ Note that variable names for strata and psu are VARSTR and VARPSU respectively in longitudinal files for panel 9 and beyond. These variables were named differently in the longitudinal files for panel 7 (varstrp7, varpsup7) and panel 8 (varstrp8, varpsup8) and need to be standardized when pooling with subsequent panels.

TOTOPUY1, TOTPTRY1, TOTOTHY1) on the Panel 12 longitudinal file. Thus, for 1,667 persons the values of these variables differ from those of the corresponding variables on the 2007 full year consolidated file. (The duplicate fills were also used to construct the summary measures in the 2007 and 2008 full year consolidated files. However, the duplication does not affect summary analyses of those files, because the change was implemented for both panels 12 and 13.)

Some longitudinal file users conduct analyses that require additional information about the types of drugs acquired, for example, the number of times a person obtained a specific drug (say, Lipitor) during the two-year period. These details are obtained from the 2007 and 2008 Prescription Medicine files. For Panel 12, these users will need to remove duplicate round 3 records from either the 2007 or 2008 Prescription Medicine file.

For users summarizing information over the two-year period (for example, number of Lipitor fills 2007 through 2008) or summarizing information in each round (number of Lipitor fills in rounds 1, 2, 3, 4, and 5), there is a simple method to remove duplicates. This method is also consistent with the way fills were allocated in Panels 1 through 11, so it is most appropriate when combining Panel 12 with prior panels. The variable DUP2007 identifies on the 2008 Prescription Medicine file that duplicate acquisitions on the 2007 Prescription Medicine file. Delete the 2008 records with DUP2007=1. Note, however, that deleting the 2008 records with DUP2007=1 and aggregating the remaining 2008 prescription drug records will yield year 2 use and expenditures that differ from the 2008 full year file, but the totals for year 1 will agree with the 2007 full year file at the person level.

Users summarizing information separately by year (for example, Lipitor fills in 2007 compared with Lipitor fills in 2008) or combining Panel 12 and subsequent panels need to remove duplicate fills from 2007 Prescription Medicine file rather than the 2008 Prescription Medicine file. The following steps are recommended:

1. From the 2008 Prescription Medicine file, select the records with DUP2007=1. Create a person-drug level (LINKIDX) file containing a variable counting the number of records (fills) for each LINKIDX.
2. From the 2007 Prescription Medicine file, select the panel 12 round 3 records (PANEL=12 and PURCHRD=3). Create a person-drug level (LINKIDX) file containing a variable counting the number of records (fills) for each LINKIDX.
3. Merge the information about the drugs with duplicate fills (from the 2008 data) onto the 2007 round 3 drugs by LINKIDX.
4. Reduce the 2007 round 3 fills for each drug by the number of duplicate fills.

For most person-drugs, the adjusted 2007 round 3 expenditures are the total across fills on the 2007 file minus the total across the 2008 duplicate fills. Similarly, for each source of

payment, the adjusted 2007 round 3 expenditures for a drug are the amount from the 2007 file minus the amount on the 2008 fills. This was the method used to create values of the summary expenditure variables for the most persons on the longitudinal file. Removing the duplicate expenditures is complicated, however, in two situations. (1) For a few person-drugs, total expenditures on the 2008 duplicate fills exceed the total from the 2007 round 3 records. (This occurs due to variation between the years in matching drugs reported by pharmacies to those reported by households.) (2) For some person-drugs, the sources of payment in 2008 differ from those in 2007. (This occurs due to improvements, also implemented starting with the 2008 data, in reconciling sources of payments between the pharmacy and household when the pharmacy information is imputed.) In both types of problem cases, subtracting the 2008 expenditures for the duplicate records from 2007 round 3 expenditure variables would yield negative amounts. For both types of problem cases, the method used to produce the year 1 summary variables in the longitudinal file was to reduce 2007 round 3 total expenditures on the drug in proportion to the number of fills that are duplicates; that is, adjusted 2007 round 3 total expenditures for drug = (2007 round 3 total expenditures for drug) – [(number of duplicate fills) ÷ (number of fills on the 2007 Prescription Medicine file)] × (2007 round 3 expenditures for drug). Similarly, for each payer, reduce that payer’s expenditures by the same proportion. For example, for these person-drugs, adjusted 2007 round 3 Medicare expenditures = 2007 round 3 Medicare expenditures – [(number of duplicate fills) ÷ (number of fills on the 2007 Prescription Medicine file)] × (2007 round 3 Medicare expenditures for drug).

The reallocation was made to improve the accuracy of the microdata, so the second deduplication method (removing 2007 records) is preferred in most other situations. For example, the second method may be better when users combine Panel 12 with both preceding and subsequent panels and summarize information annually. The impact of the specific method selected may be smallest for users combining Panel 12 with both preceding and subsequent panels and summarizing drug expenditures over the two-year window or by round.